

# Robust image reconstruction from multi-view measurements

Gilles Puy and Pierre Vanderghenst

**Abstract.** We propose a novel method to accurately reconstruct a set of images representing a single scene from few linear multi-view measurements. Each observed image is modeled as the sum of a background image and a foreground one. The background image is common to all observed images but undergoes geometric transformations, as the scene is observed from different viewpoints. In this paper, we assume that these geometric transformations are represented by a few parameters, e.g., translations, rotations, affine transformations, etc.. The foreground images differ from one observed image to another, and are used to model possible occlusions of the scene. The proposed reconstruction algorithm estimates jointly the images and the transformation parameters from the available multi-view measurements. The ideal solution of this multi-view imaging problem minimizes a non-convex functional, and the reconstruction technique is an alternating descent method built to minimize this functional. The convergence of the proposed algorithm is studied, and conditions under which the sequence of estimated images and parameters converges to a critical point of the non-convex functional are provided. Finally, the efficiency of the algorithm is demonstrated using numerical simulations for applications such as compressed sensing or super-resolution.

**Keywords.** inverse problem, image registration, non-convex optimization.

## 1. Introduction

Multi-view imaging has become more and more popular in the signal processing community during the past years with, for example, the design of camera arrays [1, 43] and the rise of new applications such as 3D reconstruction of a scene [19, 36].

In most multi-view imaging applications, the pose of the cameras, or of the sensing systems, is estimated by pre-calibration or by using side-information. In this paper, we study the problem of reconstructing a set of images, representing a single scene, from few measurements made at different viewpoints in absence of (or with inaccurate) prior pose estimation. This situation may occur in different scenarios.

For example, let us imagine that we have a camera that can be sent up to take pictures of the ground and that we are interested in obtaining higher resolution images from these acquisitions. Without more knowledge on the acquisition strategy, we can treat each image independently using a single-frame super-resolution technique. However, if we know that the duration between the capture of two images is not too long, it is highly probable that parts of the scene visible in one frame remain visible in the subsequent ones. Consequently, we can surely benefit from the high inter-correlation between observations to obtain better super-resolved images by jointly reconstructing several frames together, as in, e.g., [16, 41]. The inter-correlation may be modeled using geometric transformations, which depend on the camera pose, that register the images with respect to each other. In the absence of any side information, these transformations have to be estimated along with the high resolution images during the reconstruction process.

As another example, let us consider the problem of designing a similar camera as above but with strong power consumption constraints. To design such an energy-efficient system, compressed sensing (CS) is a powerful tool [12, 22]. Indeed, this theory states that sparse signals can be sampled with just a few linear and non-adaptive measurements. In scenarios where signals related to common phenomena are acquired, it is actually possible to reduce even further the number of measurements by jointly reconstructing the signals with a method exploiting the inter-correlation between them [6, 15]. A technique, as in [23, 42], able to estimate both the geometric transformations between the observed

---

This work was partly funded by the Hasler Foundation (project number 12080).

G. Puy and P. Vanderghenst are with the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. E-mail: [gilles.puy@epfl.ch](mailto:gilles.puy@epfl.ch) and [pierre.vanderghenst@epfl.ch](mailto:pierre.vanderghenst@epfl.ch).

images and the images themselves from the compressed measurements can thus help us to meet our power consumption constraints.

In the present work, we propose a novel method that estimates jointly a set of correlated images and the geometric transformations that align these images on each other. This technique also robustly handles the appearance of new objects in the scene and can be applied in different settings, such as the ones described above.

### 1.1. Problem formulation

Let  $\mathbf{u} = (u_1, u_2) \in \mathbb{R}^2$  represent the Cartesian coordinates of a point on the Euclidean plane, and  $G = \{\mathbf{u}_k\}_{1 \leq k \leq n}$  be a square uniform grid of  $n$  points. We model continuous images as functions  $x: \mathbb{R}^2 \rightarrow \mathbb{R}$  living in a space  $\{x(\mathbf{u}) = \sum_{i=1}^n x_i \varphi(\mathbf{u} - \mathbf{u}_i), x_i \in \mathbb{R}, \mathbf{u}_i \in G, i = 1, \dots, n\}$ , where  $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$  is a generating function [40]. For simplicity, we suppose that  $\varphi(\mathbf{u}_k - \mathbf{u}_i) = \delta(\mathbf{u}_k - \mathbf{u}_i)$  for all  $\mathbf{u}_k, \mathbf{u}_i \in G$ , so that the discrete image  $\mathbf{x} \in \mathbb{R}^n$  obtained by sampling  $x$  on the grid  $G$  has samples  $x_1, \dots, x_n$ .

In our setting, several observers provide different observations  $\mathbf{y}_1, \dots, \mathbf{y}_l \in \mathbb{R}^m$ , with  $m \leq n$ , of a scene from different viewpoints. As a first approximation, one can consider that these observations describe a single image  $x_0$ , referred as the *background* image hereafter. This image is “viewed” from different perspectives and thus undergoes geometric transformations. We consider here that these transformations are unknown and need to be estimated. However, we restrict ourselves to global transformations represented by few parameters, such as translations, rotations, or affine transformations. For each observer  $j$ , with  $j = 1, \dots, l$ , these transformations are modeled using a function  $\tau_{\theta_j}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , depending on  $q$  parameters  $\theta_j \in \mathbb{R}^q$ , that maps the coordinates  $\mathbf{u}$  into new coordinates  $\tau_{\theta_j}(\mathbf{u})$ . The background image transformed by  $\tau_{\theta_j}$  is thus  $x_0 \circ \tau_{\theta_j}$ .

To complete the model, we also take into account possible occlusions of the scene. These occlusions may obviously be different from one observer to another. We model them here using  $l$  *foreground* images  $x_1, \dots, x_l$  and write the  $j^{\text{th}}$  observed image as  $x_0 \circ \tau_{\theta_j} + x_j$ .

Finally, we assume that the observations  $\mathbf{y}_j$  are obtained by linear projection of  $x_0 \circ \tau_{\theta_j} + x_j$  onto  $m$  known functions  $a_1^j, \dots, a_m^j: \mathbb{R}^2 \rightarrow \mathbb{R}$ . These functions model the acquisition system with which the images are observed. They can represent a blurring operator for a deconvolution problem, or random waveforms in a compressed sensing setting. The  $i^{\text{th}}$  entry of  $\mathbf{y}_j$  satisfies

$$\begin{aligned} y_{ij} &= \int_{\mathbb{R}^2} (x_0 \circ \tau_{\theta_j}(\mathbf{u}) + x_j(\mathbf{u})) a_i^j(\mathbf{u}) d\mathbf{u} \\ &= \int_{\mathbb{R}^2} \sum_{k=1}^n (x_k^0 \varphi(\tau_{\theta_j}(\mathbf{u}) - \mathbf{u}_k) + x_k^j \varphi(\mathbf{u} - \mathbf{u}_k)) a_i^j(\mathbf{u}) d\mathbf{u}, \end{aligned}$$

where  $x_k^0$  and  $x_k^j$ , with  $k = 1, \dots, n$ , are the samples describing  $x_0$  and  $x_j$  respectively. To facilitate the implementation of this observation model, we assume that the grid  $G$  has a sufficiently high resolution so that the integral above is “well” approximated by its Riemann sum on this grid. In the following, we consider that

$$y_{ij} = \sum_{k'=1}^n \sum_{k=1}^n x_k^0 \varphi(\tau_{\theta_j}(\mathbf{u}_{k'}) - \mathbf{u}_k) a_i^j(\mathbf{u}_{k'}) + \sum_{k=1}^n x_k^j a_i^j(\mathbf{u}_k),$$

where we supposed that the pixels of the grid have a unit square area. Therefore, the measurements vector  $\mathbf{y}_j$  is equal to  $\mathbf{A}_j(\mathbf{S}(\theta_j)\mathbf{x}_0 + \mathbf{x}_j)$ , where  $\mathbf{A}_j = (a_i^j(\mathbf{u}_k))_{ik} \in \mathbb{R}^{m \times n}$ , and  $\mathbf{S}: \mathbb{R}^q \rightarrow \mathbb{R}^{n \times n}$  satisfies

$$(1.1) \quad \mathbf{S}(\theta_j) = \begin{bmatrix} \varphi(\tau_{\theta_j}(\mathbf{u}_1) - \mathbf{u}_1) & \dots & \varphi(\tau_{\theta_j}(\mathbf{u}_1) - \mathbf{u}_n) \\ \vdots & & \vdots \\ \varphi(\tau_{\theta_j}(\mathbf{u}_n) - \mathbf{u}_1) & \dots & \varphi(\tau_{\theta_j}(\mathbf{u}_n) - \mathbf{u}_n) \end{bmatrix}.$$

Concatenating all the observations in a single vector, we have

$$(1.2) \quad \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_l \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \mathbf{S}(\boldsymbol{\theta}_1) & \mathbf{A}_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_l \mathbf{S}(\boldsymbol{\theta}_l) & 0 & \dots & \mathbf{A}_l \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_l \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_l \end{bmatrix},$$

where we introduced  $l$  vectors  $\mathbf{n}_1, \dots, \mathbf{n}_l \in \mathbb{R}^m$  to model additive measurement noise. Our goal is to design a method that estimates the discrete images  $\mathbf{x} = (\mathbf{x}_0^\top, \dots, \mathbf{x}_l^\top)^\top \in \mathbb{R}^{(l+1)n}$ , and the transformation parameters  $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_l^\top)^\top \in \mathbb{R}^{lq}$ , using  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_l^\top)^\top \in \mathbb{R}^{lm}$  as sole information.

The inter-correlation between the observed images appears explicitly in the measurement model (1.2). We consider that the observed images share a common signal - the background image - which lives on a low-dimensional manifold. In the literature, other models that exploit the inter-correlation between the measurements have been proposed.

## 1.2. Related works

The problem of reconstructing an image from multi-view measurements has been studied by several authors in contexts such as compressed sensing, super-resolution, or robust image alignment. We try here to give a brief overview of the different types of techniques used.

When the measurements are obtained by compressed sensing, some techniques, like ours, rely on the estimation of geometric transformations between images. In [13], the authors propose to reconstruct the background image using an over-complete dictionary created by scaling, translating and rotating several times a mother waveform. The image is assumed to have a sparse decomposition in this dictionary, and the reconstruction is done by using the observations  $\mathbf{y}_1, \dots, \mathbf{y}_l$  one after another. We note that this method does not deal with the problem of occlusions, and that, as mentioned by the authors, the quality of the final reconstruction depends on the order in which the measurements are taken. In [23, 42], the authors suppose that the observed images live exactly on a low-dimensional manifold. Let us remark that it would be the case in our model in absence of occlusions. For the reconstruction, the problem is separated into two separate tasks: registration and image estimation. The authors use a manifold lifting algorithm for the registration and standard compressed sensing algorithms for the image estimation. We note that to work efficiently, the manifold lifting algorithm requires the use of a specific measurement matrix, implemented with noiselets, to estimate the transformation parameters in a coarse-to-fine scales fashion.

Continuing with compressed sensing, other techniques rely on more general interpolation model between images. In [37, 38], the images are first estimated with a bidirectional interpolation using two known neighbor images and then corrected using the acquired measurements. An iterative process, alternating between interpolation and reconstruction, is then proposed to refine the estimation of the images. Note that a similar idea is used in [2] for dynamic MRI. In [21], a slightly simpler model is used. The authors assume that the difference between two images is sparse in a wavelet basis and that the background image is sparse in the gradient domain. Then, a convex minimization problem is solved to reconstruct jointly all the images. In [14, 17], the interpolation relies on “disparity maps” that indicate the correspondences between the pixels of two images. These maps are used to estimate the images from the measurements, and the newly estimated pictures can be used to refine the disparity maps. The process is then iterated several times. Let us remark that the authors do not provide any proof of convergence. We note that the authors deal with the problem of occlusions by supposing that they are sparse in image space. Let us also mention that a similar idea is proposed in [35]. Finally, in [30], the authors use a linear dynamical system to model the intercorrelation between frames, and reconstruct a video sequence.

When the matrices  $\mathbf{A}_1, \dots, \mathbf{A}_m$  implement a blurring and downsampling operator, one can try to obtain a high-resolution background image from the images observed at low-resolution by solving (1.2). Super-resolution from multiple frames has also been widely studied and several image priors and correlation models between images have been proposed to solve this problem. As before, some techniques model the correlation between images using geometric transformations [16, 41] while other ones use more general correlation models [26, 32].

Finally, let us highlight a similarity between our model and the one used in [24, 25] for robust image alignment, which consists in aligning a set of images despite the presence of large occlusions. With the measurement model (1.2), one can attempt to solve this problem by initializing the vectors  $\mathbf{y}_1, \dots, \mathbf{y}_l$  with the observed images and the observations matrices  $\mathbf{A}_1, \dots, \mathbf{A}_l$  with the identity. Then, the algorithm should estimate the background image, the transformation parameters that align this background image on the observed ones, and the foreground images modeling occlusions. In [24, 25], a different approach is proposed. The authors concatenate the initial images  $\mathbf{y}_1, \dots, \mathbf{y}_l$  in a matrix and explain that, after alignment of these images, this matrix can be decomposed as the sum of a low-rank matrix (modeling correlated components) and a sparse matrix (modeling occlusions). This model serves as the criterion to align the images. Note that in the limit where the rank of the final matrix modeling the correlated components is 1, their method, like ours, also decomposes the initial images into a single background image and  $l$  foreground images.

### 1.3. Main contribution and organization of the paper

We propose here a novel algorithm for *joint* registration and reconstruction of a set of misaligned images, and, unlike the related methods presented in the previous section, study the *convergence* of the algorithm. Our method relies on the measurement model (1.2), and is robust to measurement noise and occlusions. The technique can be applied in different scenarios such as compressed sensing, super-resolution, or robust image registration.

In Section 2, we identify the solution of the multi-view imaging problem with the global minimizer of a non-convex functional, and propose an algorithm for the minimization of this functional. In Section 3, we study the convergence of the reconstruction method, and prove that the sequence of *estimates* (images and transformation parameters) converges to a critical point of the functional. The efficiency of the algorithm is then demonstrated experimentally in Section 4 for several problems. In the spirit of reproducible research, our code<sup>1</sup> and data needed to reproduce the results presented in this paper can be downloaded openly at <http://lts2www.epfl.ch/people/gilles/software>. We finally conclude in Section 5. Some results required to prove the convergence of the algorithm are gathered in Appendix.

Note that we already briefly studied the problem of reconstructing a set of images from multi-view measurements in [27], where the measurement model (1.2) was also used. However, the technique proposed to reconstruct the images was based on the Bregman iterative regularization procedure, which is different from the one used here. Furthermore, the convergence results presented in [27] essentially concern the decrease of the objective value and, unlike here, do not ensure the convergence of the sequence of estimated images and parameters.

### 1.4. Notations and definitions

The Euclidean scalar product of  $\mathbb{R}^n$  is denoted  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|_2$  is the corresponding  $\ell_2$ -norm. The  $\ell_1$ -norm of a vector  $\mathbf{x} = (x_i)_{1 \leq i \leq n} \in \mathbb{R}^n$  is defined as  $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ .

Let  $C$  be a closed convex subset of  $\mathbb{R}^n$ . The indicator function of  $C$  is denoted  $i_C: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ . It is a proper lower semicontinuous convex function that satisfies

$$i_C(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in C, \\ +\infty, & \text{otherwise.} \end{cases}$$

Let  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function. The domain of  $f$  is denoted and defined by  $\text{dom} f = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < +\infty\}$ . The subdifferential of  $f$  at  $\mathbf{x} \in \text{dom} f$  is denoted by  $\partial f(\mathbf{x})$  (see, e.g., Section 2 of [3] for the definition). Note that  $\partial f(\mathbf{x}) = \emptyset$  if  $\mathbf{x} \notin \text{dom} f$ . A necessary (but not sufficient) condition for  $\mathbf{x}^* \in \mathbb{R}^n$  to be a minimizer of  $f$  is  $\partial f(\mathbf{x}^*) \ni \mathbf{0}$ . A point that satisfies the last condition is called a critical point of  $f$ .

Finally,  $\mathcal{C}^k$  denotes the set of functions that are  $k$  times continuously differentiable.

---

<sup>1</sup>We used the SPARCO toolbox available at <http://www.cs.ubc.ca/labs/sc1/sparco/> [9].

## 2. A non-convex optimization technique

In this section, we first identify the solution of the compressed multi-view imaging problem with the minimizer of a non-convex functional and propose an algorithm to minimize this functional.

### 2.1. Solution as minimizer of a non-convex functional

Let  $\epsilon \geq 0$  be an upper bound on the energy of the measurement noise, i.e.,  $\|\mathbf{n}\|_2 \leq \epsilon$ . Our objective is to find a set of parameters  $\boldsymbol{\theta}^*$  and images  $\mathbf{x}^*$  that satisfy the constraint  $\|\mathbf{A}(\boldsymbol{\theta}^*) \mathbf{x}^* - \mathbf{y}\|_2 \leq \epsilon$ , where

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{A}_1 \mathbf{S}(\boldsymbol{\theta}_1) & \mathbf{A}_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_l \mathbf{S}(\boldsymbol{\theta}_l) & 0 & \dots & \mathbf{A}_l \end{bmatrix} \in \mathbb{R}^{lm \times (l+1)n}.$$

As infinitely many solutions satisfy this constraint, prior information is needed to restrict the set of admissible solutions. To regularize this ill-posed inverse problem, we can, for example, search for images sparse in a wavelet basis by minimizing the  $\ell_1$ -norm of their wavelet coefficients, or seek a solution that minimizes the Total Variation (TV) norm if the images are piecewise constant [29]. We may also require the transformation parameters  $\boldsymbol{\theta}_j$ , with  $j = 1, \dots, l$ , to belong to some convex sets  $\Theta_j = \{\boldsymbol{\theta}_j \in \mathbb{R}^q : \underline{\boldsymbol{\theta}}_j \leq \boldsymbol{\theta}_j \leq \bar{\boldsymbol{\theta}}_j\}$  where  $\underline{\boldsymbol{\theta}}_j \in \mathbb{R}^q$  and  $\bar{\boldsymbol{\theta}}_j \in \mathbb{R}^q$  are pre-defined upper and lower bounds<sup>2</sup>. Therefore, an estimate of the images and transformation parameters can be obtained by solving a minimization problem of the form

$$(2.3) \quad \min_{(\mathbf{x}, \boldsymbol{\theta})} f(\mathbf{x}) + \kappa \|\mathbf{A}(\boldsymbol{\theta}) \mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to } \boldsymbol{\theta} \in \Theta,$$

where  $f: \mathbb{R}^{(l+1)n} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower-semicontinuous convex function,  $\kappa > 0$  is a regularizing parameter that should be adjusted with the noise level  $\epsilon$ , and  $\Theta = \{\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_l^\top)^\top \in \mathbb{R}^{lq} : \boldsymbol{\theta}_j \in \Theta_j, j = 1, \dots, l\}$ .

Problem (2.3) is, in general, non-convex and finding a global minimizer is not trivial. Nevertheless, one can still try to find a local minimum using an alternating minimization technique such as the block coordinate descent method [39]. Recently, Attouch *et al.* proposed in [3] a new type of alternating algorithm for the minimization of a non-convex functional and, more importantly, managed to prove that the sequence generated by the algorithm converges to a critical point of the functional for a wide class of problems. As said by the authors, their algorithm can be interpreted as a proximal regularization of the Gauss-Seidel method where cost-to-move functions are added in the minimization procedure. These convergence results were then generalized to inexact descent methods that satisfy a sufficient decrease condition [4]. Based on this work, we develop a minimization method for problem (2.3) and prove that the generated sequence converges to a critical point  $(\mathbf{x}^*, \boldsymbol{\theta}^*)$  of the functional  $L: \mathbb{R}^{(l+1)n \times lq} \rightarrow \mathbb{R} \cup \{+\infty\}$  defined as

$$(2.4) \quad L(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}) + \kappa \|\mathbf{A}(\boldsymbol{\theta}) \mathbf{x} - \mathbf{y}\|_2^2 + i_\Theta(\boldsymbol{\theta}).$$

We recall that this condition is necessary for  $(\mathbf{x}^*, \boldsymbol{\theta}^*)$  to be a global minimizer of Problem (2.3) but is, unfortunately, not sufficient.

The proposed algorithm is a descend method, i.e., it generates a sequence of estimates  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  such that  $L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1}) \leq L(\mathbf{x}^k, \boldsymbol{\theta}^k)$  for all  $k \in \mathbb{N}$ , and consists of two steps. We start by describing each of these steps and study the convergence of the algorithm in Section 3.

### 2.2. First step of the algorithm

Let  $(\mathbf{x}^k, \boldsymbol{\theta}^k) \in \mathbb{R}^{(l+1)n} \times \Theta$  be the estimates obtained after  $k$  iterations of the algorithm. The first step consists in finding a new estimate  $\mathbf{x}^{k+1} \in \mathbb{R}^{(l+1)n}$  that decreases the value of the objective function  $L$  while keeping  $\boldsymbol{\theta}^k$  fixed. Let us choose  $\mathbf{x}^{k+1}$  as a solution of

$$(2.5) \quad \min_{\mathbf{x} \in \mathbb{R}^{(l+1)n}} f(\mathbf{x}) + \kappa \|\mathbf{A}(\boldsymbol{\theta}^k) \mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda_{\mathbf{x}}}{2} g(\mathbf{x} - \mathbf{x}^k),$$

<sup>2</sup>Let  $\bar{\boldsymbol{\theta}} = (\bar{\theta}_i)_{1 \leq i \leq q} \in \mathbb{R}^q$ ,  $\boldsymbol{\theta} = (\theta_i)_{1 \leq i \leq q} \in \mathbb{R}^q$ ,  $\boldsymbol{\theta} \leq \bar{\boldsymbol{\theta}}$  means that  $\theta_i \leq \bar{\theta}_i$  for all  $i \in \{1, \dots, q\}$ .

where  $\lambda_{\mathbf{x}} > 0$ , and  $g: \mathbb{R}^{(l+1)n} \rightarrow \mathbb{R}$  is proper lower-semicontinuous convex function such that  $g(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^{(l+1)n}$ . It is clear that (2.5) implies

$$(2.6) \quad L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k) + \frac{\lambda_{\mathbf{x}}}{2} g(\mathbf{x}^{k+1} - \mathbf{x}^k) \leq L(\mathbf{x}^k, \boldsymbol{\theta}^k).$$

Hence  $L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k) \leq L(\mathbf{x}^k, \boldsymbol{\theta}^k)$  with a decrease of  $\lambda_{\mathbf{x}} g(\mathbf{x}^{k+1} - \mathbf{x}^k)/2$ , at least. Note that the above minimization problem is convex in  $\mathbf{x}$  and that it can be solved using, e.g., the algorithms presented in [8].

In Problem (2.5), the function  $g$  acts as a proximal term and  $\lambda_{\mathbf{x}}$  as a stepsize. It ensures a sufficient decrease of the functional  $L$  at each iteration and is essential to prove the convergence of the sequence  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  to a critical point. This function also provides, to some extent, a control on the evolution of the generated sequence, and we realized that it must be chosen carefully to achieve good reconstructions. Note that in the work [3] of Attouch *et al.*, the function  $g$  is a squared  $\ell_2$ -norm.

From the multi-view measurements  $\mathbf{y}_1, \dots, \mathbf{y}_l$ , the algorithm should reconstruct the background image, separate the occlusions (foreground), and provide the transformation parameters that register the initial images between them. This problem is solved here by alternatively refining the estimation of the images  $\mathbf{x}$  (background and foreground) and of the transformation parameters  $\boldsymbol{\theta}$  (image registration). We noticed that to improve the quality of the registration, it is better to have a reconstruction of the images in a coarse-to-fine scales fashion, as in [23, 42]. In these papers, the authors use a dedicated measurement matrix that allows this type of reconstruction. In this paper, this goal is achieved by choosing a function  $g$  that favors the reconstruction of the coarse scales first and then of the fine scales. There is thus no constraint on the choice of the measurement matrices anymore.

In order to have a coarse-to-fine scales reconstruction of the images, we use a wavelet tight-frame  $\mathbf{W} \in \mathbb{R}^{n \times p}$ , with  $p \geq n$ , i.e.,  $\mathbf{W}\mathbf{W}^\top$  is equal to  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ , the identity matrix. One should remark that most of the largest wavelet coefficients of a natural image live at the coarsest scales and that the amplitude of these coefficients decreases when going to finer scales (except at the few singularities). To achieve our goal, we need a function  $g$  that favors a reconstruction of the coarse scales first, by selecting the biggest wavelet coefficients, and then of the fine scales, by selecting the smallest wavelet coefficients. Let  $\boldsymbol{\Psi} \in \mathbb{R}^{(l+1)n \times (l+1)p}$  be the block-diagonal matrix built by repeating  $l+1$  times the matrix  $\mathbf{W}$  on the diagonal. We have  $\boldsymbol{\Psi}\boldsymbol{\Psi}^\top = \mathbf{I}_{(l+1)n}$ . Take  $\mathbf{x}^{k+1}$  as a minimizer of  $f(\mathbf{x}) + \kappa \|\mathbf{A}(\boldsymbol{\theta}^k) \mathbf{x} - \mathbf{y}\|_2^2 + \lambda_{\mathbf{x}} \|\boldsymbol{\Psi}^\top(\mathbf{x} - \mathbf{x}^k)\|_1/2$ , and remember that the  $\ell_1$ -norm favors solutions with few large coefficients. Hence, the function  $f$  forces  $\mathbf{x}^{k+1}$  to fit our prior, the quadratic term forces  $\mathbf{x}^{k+1}$  to be consistent with the observations, and the cost-to-move function  $\mathbf{x} \mapsto \|\boldsymbol{\Psi}^\top(\mathbf{x} - \mathbf{x}^k)\|_1$  imposes that  $\mathbf{x}^{k+1}$  differs from  $\mathbf{x}^k$  by only a few large wavelet coefficients. Starting from  $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^{(l+1)n}$  at  $k=0$ , the first estimated images  $\mathbf{x}^1$  have thus a sparse decomposition in the wavelet frame  $\mathbf{W}$ . The sparsity increases with the stepsize parameter  $\lambda_{\mathbf{x}}$ . As we are reconstructing natural images, these few large coefficients mainly live at large scales and  $\mathbf{x}^1$  is a coarse approximation of the solution. The estimation of the images is then refined at the next iteration, and the cost-to-move function favors the selection of the next biggest wavelet coefficients living at finer scales. We thus have a coarse-to-fine scales reconstruction of the images. This behavior will be illustrated in Section 4.

Note that to be able to prove the convergence of the sequence  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  to a critical point of  $L$  with the results presented in [3, 4], we actually substitute the Huber function for the  $\ell_1$ -norm as cost-to-move function. The Huber function is a smooth approximation of the  $\ell_1$ -norm that depends on a smoothing parameter  $\mu > 0$ . This parameter can be chosen small so that both functions are nearly identical. Let  $\boldsymbol{\alpha} = (\alpha_i)_{1 \leq i \leq p} \in \mathbb{R}^{(l+1)p}$ , the Huber function  $h_\mu: \mathbb{R}^{(l+1)p} \rightarrow \mathbb{R}$  satisfies

$$h_\mu(\boldsymbol{\alpha}) = \sum_{i=1}^{(l+1)p} h_i,$$

where

$$h_i = \begin{cases} \alpha_i^2/(2\mu), & \text{if } |\alpha_i| < \mu, \\ |\alpha_i| + \mu/2, & \text{otherwise,} \end{cases} \quad \forall i \in \{1, \dots, (l+1)p\}.$$

From now on, the term  $g(\mathbf{x} - \mathbf{x}^k)$  in Problem (2.5) is replaced by  $h_\mu(\boldsymbol{\Psi}^\top(\mathbf{x} - \mathbf{x}^k))$ .

### 2.3. Second step of the algorithm

The second step consists in updating the transformation parameters to further decrease the value of the objective function  $L$ . As the functions  $\theta \mapsto \|A(\theta)x - y\|_2^2$  and  $i_\Theta$  are separable in  $\theta_j$ , with  $j = 1, \dots, l$ , we optimize the transformation parameters separately for each observations.

Let  $(x^k, \theta^k) \in \mathbb{R}^{(l+1)n} \times \Theta$  be the estimates obtained after  $k$  iterations of the algorithm and  $x^{k+1} \in \mathbb{R}^{(l+1)n}$  be the solution of problem (2.5). To simplify the notations, we introduce new functions  $Q_j^{k+1}: \mathbb{R}^q \rightarrow \mathbb{R}$ , with  $j = 1, \dots, l$ , defined as

$$(2.7) \quad Q_j^{k+1}(\theta_j) = \|A_j S(\theta_j) x_0^{k+1} + A_j x_j^{k+1} - y\|_2^2.$$

Note that  $\|A(\theta)x^{k+1} - y\|_2^2 = \sum_{j=1}^l Q_j^{k+1}(\theta_j)$  and that  $i_\Theta(\theta) = \sum_{j=1}^l i_{\Theta_j}(\theta_j)$ . We choose to update the transformation parameters with the following projected Newton-like method [10, 31].

Let us assume that the entries of the matrix  $S(\theta_j)$  are differentiable with respect to the transformation parameters. The first order Taylor expansion of  $S(\theta_j)x_0^{k+1}$  at  $\theta_j^k$  is  $S(\theta_j^k)x_0^{k+1} + J(\theta_j^k)(\theta_j - \theta_j^k)$  with

$$J(\theta_j^k) = (\partial_{\theta_{1j}} S(\theta_j^k) x_0^{k+1}, \dots, \partial_{\theta_{qj}} S(\theta_j^k) x_0^{k+1}) \in \mathbb{R}^{n \times q}.$$

Therefore,

$$Q_j^{k+1}(\theta_j^k) + \langle \nabla Q_j^{k+1}(\theta_j^k), \theta_j - \theta_j^k \rangle + \|A_j J(\theta_j^k)(\theta_j - \theta_j^k)\|_2^2,$$

with

$$\nabla Q_j^{k+1}(\theta_j^k) = 2 (A_j J(\theta_j^k))^T (A_j S(\theta_j^k) x_0^{k+1} + A_j x_j^{k+1} - y),$$

is a second order approximation of  $Q_j^{k+1}$  at  $\theta_j^k$ . The positive semi-definite matrix

$$(2.8) \quad H(\theta_j^k) = 2 (A_j J(\theta_j^k))^T (A_j J(\theta_j^k)),$$

can be viewed as an approximation of the Hessian of the function  $Q_j^{k+1}$ . To update the transformation parameters, we chose to minimize this quadratic approximation to which we add another quadratic term that ensures a decrease of the objective function. We take as next estimate of the transformation parameters

$$(2.9) \quad \theta_j^{k+1} = \underset{\theta_j \in \Theta_j}{\operatorname{argmin}} \langle \nabla Q_j^{k+1}(\theta_j^k), \theta_j - \theta_j^k \rangle + \frac{1}{2} \langle \theta_j - \theta_j^k, [H(\theta_j^k) + 2^i \lambda_\theta I_q] (\theta_j - \theta_j^k) \rangle,$$

where  $I_q \in \mathbb{R}^{q \times q}$  is the identity matrix,  $\lambda_\theta > 0$ , and  $i$  is the smallest positive integer such that

$$(2.10) \quad \begin{aligned} Q_j^{k+1}(\theta_j^{k+1}) &\leq Q_j^{k+1}(\theta_j^k) + \langle \nabla Q_j^{k+1}(\theta_j^k), \theta_j^{k+1} - \theta_j^k \rangle \\ &\quad + \frac{1}{2} \langle \theta_j^{k+1} - \theta_j^k, [H(\theta_j^k) + (2^i - 1) \lambda_\theta I_q] (\theta_j^{k+1} - \theta_j^k) \rangle. \end{aligned}$$

The existence of such an  $i$  is discussed in the next section. Let us remark that if the minimization (2.9) was performed over all  $\mathbb{R}^q$  instead of  $\Theta_j$ , the solution would be  $\theta_j^{k+1} = \theta_j^k - [H(\theta_j^k) + 2^i \lambda_\theta I_q]^{-1} \nabla Q_j^{k+1}(\theta_j^k)$ , as in a Newton method with Hessian  $H(\theta_j^k) + 2^i \lambda_\theta I_q$ . One can also note that if we had chosen  $H(\theta_j^k) = \mathbf{0}$  then  $\theta_j^{k+1}$  would be obtained with a simple projected gradient update. We noticed however that the Newton-like update (2.9), which can, for example, be solved using the algorithms presented in [8], yields more accurate results.

To check that our choice of new transformation parameters decreases the value of the objective function  $L$ , one just has to combine (2.9) and (2.10) to realize that

$$\begin{aligned} Q_j^{k+1}(\theta_j^{k+1}) + \frac{\lambda_\theta}{2} \|\theta_j^{k+1} - \theta_j^k\|_2^2 &\leq \\ Q_j^{k+1}(\theta_j^k) + \langle \nabla Q_j^{k+1}(\theta_j^k), \theta_j - \theta_j^k \rangle + \frac{1}{2} \langle \theta_j - \theta_j^k, [H(\theta_j^k) + 2^i \lambda_\theta I_q] (\theta_j - \theta_j^k) \rangle, \end{aligned}$$

for any  $\theta_j \in \Theta_j$ . Choosing  $\theta_j = \theta_j^k$  in the last inequality, multiplying it by  $\kappa$ , and summing all the inequalities obtained for  $j$  from 1 to  $l$  yields

$$(2.11) \quad L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1}) + \frac{\kappa \lambda_{\boldsymbol{\theta}}}{2} \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2^2 \leq L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k).$$

Therefore  $L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1}) \leq L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k)$  with a decrease of  $\kappa \lambda_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2^2 / 2$ , at least.

### 3. Convergence analysis

In this section, we analyze the convergence of the optimization method described in Section 2. We first present sufficient conditions under which the generated sequence converges to a critical point of  $L$ . Then, we describe several cases where these conditions are satisfied. In particular, we study different types of geometric transformations which meet these requirements. Finally, we discuss the influence of the stepsize parameters on the final reconstruction quality.

---

#### Algorithm 1

---

**Inputs:** measurements  $\mathbf{y} \in \mathbb{R}^{lm}$ , tight frame  $\boldsymbol{\Psi} \in \mathbb{R}^{(l+1)n \times (l+1)p}$ , regularization parameter  $\kappa > 0$ , stepsizes  $(\lambda_{\mathbf{x}}^k)_{k \in \mathbb{N}}$ ,  $\lambda_{\boldsymbol{\theta}} > 0$ , and bounds  $\bar{\boldsymbol{\theta}} \geq \underline{\boldsymbol{\theta}}$ .

**Initializations:** set  $k = 0$ ,  $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^{(l+1)n}$ , and  $\boldsymbol{\theta}^0 \in \Theta$ .

**repeat**

1) Set

$$\mathbf{x}^{k+1} \leftarrow \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^{(l+1)n}} L(\mathbf{x}, \boldsymbol{\theta}^k) + \frac{\lambda_{\mathbf{x}}^k}{2} h_{\mu}(\boldsymbol{\Psi}^{\top}(\mathbf{x} - \mathbf{x}^k)).$$

2) For all  $j = 1, \dots, l$ , set

$$\boldsymbol{\theta}_j^{k+1} \leftarrow \operatorname{argmin}_{\boldsymbol{\theta}_j \in \Theta_j} \langle \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^k), \boldsymbol{\theta}_j - \boldsymbol{\theta}_j^k \rangle + \frac{1}{2} \langle \boldsymbol{\theta}_j - \boldsymbol{\theta}_j^k, [\mathbf{H}(\boldsymbol{\theta}_j^k) + 2^i \lambda_{\boldsymbol{\theta}} \mathbf{I}] (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^k) \rangle,$$

where  $Q_j^{k+1}$  is defined in (2.7),  $\mathbf{H}(\boldsymbol{\theta}_j^k)$  is defined in (2.8), and  $i$  is the smallest positive integer such that (2.10) is satisfied.

3)  $k \leftarrow k + 1$

**until** convergence or  $k \geq k_{\max}$

**Outputs:** Estimated images  $\mathbf{x}^*$  and transformation parameters  $\boldsymbol{\theta}^*$ .

---

#### 3.1. General convergence result

Let us consider Algorithm 1 which is a summary of the optimization method described in the previous section. We are now in position to state our main convergence result.

**Theorem 1.** *Let  $L$  be the objective function defined in (2.4) with  $\kappa > 0$ . Assume that  $L$  is bounded below, that the entries of  $\mathbf{S}_j$ , with  $j = 1, \dots, l$ , are twice continuously differentiable, that  $\boldsymbol{\Psi} \in \mathbb{R}^{(l+1)n \times (l+1)p}$  satisfies  $\boldsymbol{\Psi} \boldsymbol{\Psi}^{\top} = \mathbf{I}_{(l+1)n}$ , and that the stepsizes satisfy  $0 < \underline{\lambda} \leq \lambda_{\mathbf{x}}^k, \lambda_{\boldsymbol{\theta}} \leq \bar{\lambda}$  for all  $k \in \mathbb{N}$ . Then, the sequence of estimates  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  generated by Algorithm 1 is correctly defined and the following statements hold.*

1) For all  $k \geq 0$ ,

$$(3.12) \quad L(\mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1}) + \frac{\lambda}{2} \left[ \kappa \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2^2 + h_{\mu}(\boldsymbol{\Psi}^{\top}(\mathbf{x}^{k+1} - \mathbf{x}^k)) \right] \leq L(\mathbf{x}^k, \boldsymbol{\theta}^k).$$

Hence  $L(\mathbf{x}^k, \boldsymbol{\theta}^k)$  does not increase.

2) The sequences  $(\mathbf{x}^{k+1} - \mathbf{x}^k)_{k \in \mathbb{N}}$  and  $(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  converge. Indeed,

$$(3.13) \quad \lim_{k \rightarrow +\infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 + \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2 = 0.$$



- 3) Assume that  $L$  has the Kurdyca-Łojasiewicz property (see Definition 3.2 in [3]). Then, if the sequence  $(\mathbf{x}^k)_{k \in \mathbb{N}}$  is bounded, the sequence  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  converges to a critical point  $(\mathbf{x}^*, \boldsymbol{\theta}^*)$  of  $L$ .

*Proof.* Let us start by showing that the sequence generated by Algorithm 1 is well-defined. For the first step of the algorithm, note that, for all  $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}}) \in \mathbb{R}^{(l+1)n} \times \mathbb{R}^{lq}$  and  $\tilde{\lambda} > 0$ ,  $\mathbf{x} \mapsto L(\mathbf{x}, \tilde{\boldsymbol{\theta}})$  is a proper lower semicontinuous convex function which is bounded below and that  $\mathbf{x} \mapsto \tilde{\lambda} h_\mu(\Psi^\top(\mathbf{x} - \tilde{\mathbf{x}}))/2$  is a proper lower semicontinuous convex function which is coercive. Therefore, the function  $\mathbf{x} \mapsto L(\mathbf{x}, \tilde{\boldsymbol{\theta}}) + \tilde{\lambda} h_\mu(\Psi^\top(\mathbf{x} - \tilde{\mathbf{x}}))/2$  has a minimizer (Corollary 11.15, [7]). For the second step of the algorithm, remark that, for all  $(\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_j) \in \mathbb{R}^n \times \mathbb{R}^n$ , the function  $Q_j: \boldsymbol{\theta}_j \mapsto \|\mathbf{A}_j \mathbf{S}(\boldsymbol{\theta}_j) \tilde{\mathbf{x}}_0 + \mathbf{A}_j \tilde{\mathbf{x}}_j - \mathbf{y}_j\|_2^2$  is  $\mathcal{C}^1$  with Lipschitz continuous gradient on the closed and bounded convex set  $\Theta_j$ . Let  $\Lambda_j$  be this Lipschitz constant, whose value *a priori* depends on  $(\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_j)$ . Using the descent lemma (Lemma 3.1, [4]), we have

$$Q_j(\boldsymbol{\theta}_j^1) \leq Q_j(\boldsymbol{\theta}_j^2) + \langle \nabla Q_j(\boldsymbol{\theta}_j^2), \boldsymbol{\theta}_j^1 - \boldsymbol{\theta}_j^2 \rangle + \frac{\Lambda_j}{2} \|\boldsymbol{\theta}_j^1 - \boldsymbol{\theta}_j^2\|_2^2,$$

for any two points  $(\boldsymbol{\theta}_j^1, \boldsymbol{\theta}_j^2) \in \Theta_j^2$ . This proves the existence of an integer  $i$  such that (2.10) is satisfied at each second step of Algorithm 1. The existence of a minimizer in Problem (2.9) is then proved by using, e.g., Proposition 11.14 in [7]. An induction finally shows that sequence  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  is well-defined.

Inequality (3.12) follows by combination of (2.6) and (2.11) and by using the fact that  $\lambda_{\mathbf{x}}^k, \lambda_{\boldsymbol{\theta}} \geq \underline{\lambda}$  for all  $k \in \mathbb{N}$ . Then, summing this inequality from  $k = 0$  to  $K$  yields

$$\frac{\underline{\lambda}}{2} \sum_{k=0}^K \kappa \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2^2 + h_\mu(\Psi^\top(\mathbf{x}^{k+1} - \mathbf{x}^k)) \leq L(\mathbf{x}^0, \boldsymbol{\theta}^0) - L(\mathbf{x}^{K+1}, \boldsymbol{\theta}^{K+1})$$

As  $L$  is bounded below and  $L(\mathbf{x}^0, \boldsymbol{\theta}^0) = \kappa \|\mathbf{y}\|_2^2$ , it is clear that the righthand side of the previous inequality is also bounded. Consequently,

$$\sum_{k=0}^{+\infty} h_\mu(\Psi^\top(\mathbf{x}^{k+1} - \mathbf{x}^k)) + \kappa \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2^2 < +\infty,$$

and, using the definition of  $h_\mu$  and the fact that  $\Psi\Psi^\top = \mathbf{I}_{(l+1)n}$ , the second point of the theorem holds.

The proof of the third point make use of results established by Attouch *et al.* in [4] and can be found in Appendix.  $\square$

The third point of Theorem 1 applies if  $L$  has the Kurdyca-Łojasiewicz property. As explained in [3], this property is satisfied by several types of functions including semi-algebraic ones. A function  $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is semi-algebraic if its graph  $\{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} : g(\mathbf{x}) = t\}$  is a semi-algebraic subset of  $\mathbb{R}^{n+1}$ , i.e., if it can be written as a finite union of sets of the form  $\{\mathbf{x} \in \mathbb{R}^{n+1} : p_i(\mathbf{x}) = 0, q_i(\mathbf{x}) < 0, i = 1, \dots, r\}$ , where  $p_i$  and  $q_i$  are polynomials. The  $\ell_1$ -norm is thus an example of such a function. Note that the indicator function of a semi-algebraic set is semi-algebraic, and that the set of semi-algebraic functions is stable under basics operations such as sums, products or compositions (Section 4.3, [3]).

In the definition of the objective function  $L$ , the set  $\Theta$  is semi-algebraic. Consequently,  $L$  satisfies the Kurdyca-Łojasiewicz property if, e.g.,  $f$  and the fidelity term  $\|\mathbf{A}(\boldsymbol{\theta})\mathbf{x} - \mathbf{y}\|_2^2$  are semi-algebraic. For  $f$ , one can use, for example,  $f(\mathbf{x}) = \|\Phi\mathbf{x}\|_1$  with any real matrix  $\Phi$ . This choice ensures that  $L$  is bounded below. Also note that if  $\Phi$  has full column rank, then, using condition (3.12), one can show that the sequence  $(\mathbf{x}_k)_{k \in \mathbb{N}}$  is bounded. For the fidelity term, its semi-algebraicity depends on the properties of the generating function  $\varphi$  and the type of geometric transformations. In the next section, we present examples of functions  $\varphi$  and geometric transformations such that this term is semi-algebraic and such that the entries of the matrices  $\mathbf{S}_j$ , with  $j = 1, \dots, l$ , are twice continuously differentiable, as required by Theorem 1.

### 3.2. Examples of admissible geometric transformations

Let us start by studying the simple case of translations. This type of transformation can be represented by 2 parameters  $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top \in \mathbb{R}^2$  and a function  $\tau_{\boldsymbol{\theta}}$  defined as

$$\tau_{\boldsymbol{\theta}}(\mathbf{u}) = (u_1 + \theta_1, u_2 + \theta_2)^\top.$$

Combined, for example, with the cubic interpolator defined in [18] as generating function  $\varphi(\mathbf{u}) = \phi(u_1)\phi(u_2)$  where

$$(3.14) \quad \phi(u) = \begin{cases} \frac{3}{2}|u|^3 - \frac{5}{2}|u|^2 + 1, & \text{if } 0 \leq |u| < 1, \\ -\frac{1}{2}|u|^3 + \frac{5}{2}|u|^2 - 4|u| + 2, & \text{if } 1 \leq |u| < 2, \\ 0, & \text{otherwise,} \end{cases}$$

the entries of the matrix  $S_j$  are twice continuously differentiable with respect to the transformation parameters, and the function  $(\boldsymbol{\theta}, \mathbf{x}) \mapsto \|\mathbf{A}(\boldsymbol{\theta})\mathbf{x} - \mathbf{y}\|_2^2$  is semi-algebraic. Theorem 1 thus applies in this situation. Let us remark that this choice of generating function allows us to have sparse matrices  $S_j$ , with  $j = 1, \dots, l$ , with very fast implementation. Note that, instead of the cubic interpolator, we could also have chosen any B-spline interpolators of degree larger than 3 [40].

Then, with the same choice of generating function, we can also handle the case of affine transformations. Indeed, such transformations can be represented by 6 parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_6)^\top \in \mathbb{R}^6$  with

$$(3.15) \quad \tau_{\boldsymbol{\theta}}(\mathbf{u}) = (\theta_1 u_1 + \theta_2 u_2 + \theta_3, \theta_4 u_1 + \theta_5 u_2 + \theta_6)^\top,$$

and the requirements of Theorem 1 are also met.

Next, let us show how to handle the case of homographies. An homography is usually represented using 8 parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_8)^\top \in \mathbb{R}^8$  with  $\tau_{\boldsymbol{\theta}}$  satisfying

$$\tau_{\boldsymbol{\theta}}(\mathbf{u}) = \left( \frac{\theta_1 u_1 + \theta_2 u_2 + \theta_3}{\theta_7 u_1 + \theta_8 u_2 + 1}, \frac{\theta_4 u_1 + \theta_5 u_2 + \theta_6}{\theta_7 u_1 + \theta_8 u_2 + 1} \right)^\top.$$

Unfortunately, the function  $(\boldsymbol{\theta}, \mathbf{x}) \mapsto \|\mathbf{A}(\boldsymbol{\theta})\mathbf{x} - \mathbf{y}\|_2^2$  is not semi-algebraic in this case. However, if  $|\theta_7 u_1 + \theta_8 u_2| \ll 1$ , the transformation function can be approximated by

$$(3.16) \quad \tau_{\boldsymbol{\theta}}(\mathbf{u}) = \begin{pmatrix} \theta_1 u_1 + \theta_2 u_2 + \theta_3 \\ \theta_4 u_1 + \theta_5 u_2 + \theta_6 \end{pmatrix} (1 - \theta_7 u_1 - \theta_8 u_2),$$

using a first order Taylor expansion. All the conditions are now fulfilled to apply Theorem 1. Note that the condition  $|\theta_7 u_1 + \theta_8 u_2| \ll 1$  can be enforced during the reconstruction by using appropriate bounds  $\bar{\boldsymbol{\theta}}$  and  $\underline{\boldsymbol{\theta}}$ .

Finally, let us mention that (small) rotations can be handled in the same manner by using a Taylor expansion of the sine and cosine functions.

### 3.3. Influence of the stepsizes

Let us highlight that, even though the sequence  $(\mathbf{x}^k, \boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  generated by the algorithm converges to a critical point of  $L$  for all strictly positive stepsize  $\lambda_{\boldsymbol{\theta}}$  and strictly positive bounded sequence  $(\lambda_{\mathbf{x}}^k)_{k \in \mathbb{N}}$ , different choices of stepsizes yield different results. The choice of these parameters is thus important.

For the sequence of stepsizes  $(\lambda_{\mathbf{x}}^k)_{k \in \mathbb{N}}$ , we start from large values and thus constrain  $\mathbf{x}^{k+1}$  to stay “close” to  $\mathbf{x}^k$  when the estimated images are still inaccurate. Starting with large values is also essential to have the first estimated images made of wavelets at coarse scales only. Then, we slightly decrease the stepsize value at each iteration as the estimates become more and more accurate. In all the following experiments, we use the same sequence of stepsizes:  $\lambda_{\mathbf{x}}^k = \max(0.9^k (20\kappa), 0.1)$ .

Finally, the stepsize  $\lambda_{\boldsymbol{\theta}}$  seems to have less influence on the final reconstruction quality, and we keep it fixed at 0.1 in all the experiments of the next section.

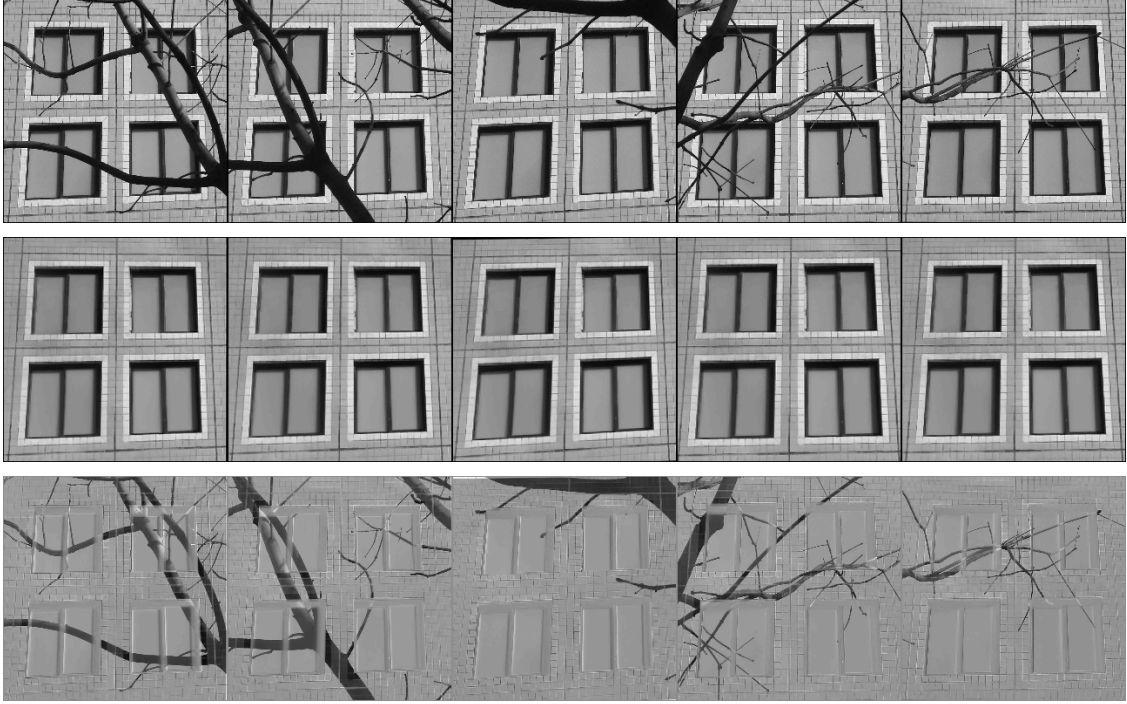


Figure 1. *Top panels:* first 5 observed images  $\mathbf{y}_1, \dots, \mathbf{y}_5$ . *Middle panels:* registered background images  $\mathbf{S}(\boldsymbol{\theta}_j^*)\mathbf{x}_0^*$ , with  $j = 1, \dots, 5$ . *Bottom panels:* estimated foreground images  $\mathbf{x}_1^*, \dots, \mathbf{x}_5^*$ .

## 4. Experiments

In this section, we demonstrate, experimentally, the efficiency of the algorithm for three different applications: robust image alignment, compressed sensing, and super-resolution. The code and data needed to reproduce the results presented in this section are available at <http://lts2www.epfl.ch/people/gilles/software>.

### 4.1. Robust image alignment

For this first experiment, we test our algorithm for robust image alignment, as studied in [24,25]. This problem consists in aligning several images of a given scene despite the presence of occlusions, as, for example, the images presented in the top panels<sup>3</sup> of Fig. 1.

In [24,25], this problem is solved with a method based on low rank and sparse approximation. To get around the difficulty of optimizing the transformations parameters that appear in their data fidelity term, the authors propose to linearize this term with respect to the transformation parameters. This approximation is valid for small transformations only but the optimization problem becomes convex and easier to solve. Then, to align images with large transformations between them, the authors propose to repeatedly solve this convex problem and linearize the data fidelity term around the new estimated parameters to improve, little by little, the registration.

Here, we solve the robust image alignment problem using the measurement model (1.2) and Algorithm 1 to minimize (2.4). In the measurement model, the matrices  $\mathbf{A}_1, \dots, \mathbf{A}_l$  are set to the identity  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ , and  $\mathbf{y}_1, \dots, \mathbf{y}_l$  contain the occluded observed images. Ideally our algorithm should estimate the background image  $\mathbf{x}_0$ , i.e., the wall and the windows in Fig. 1, extract the occlusions in  $\mathbf{x}_1, \dots, \mathbf{x}_l$ , i.e., the branches, and provide the transformation parameters  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_l$ , which align the background image on the observed images.

<sup>3</sup>Dataset available at <http://perception.csl.illinois.edu/matrix-rank/rasl.html>.

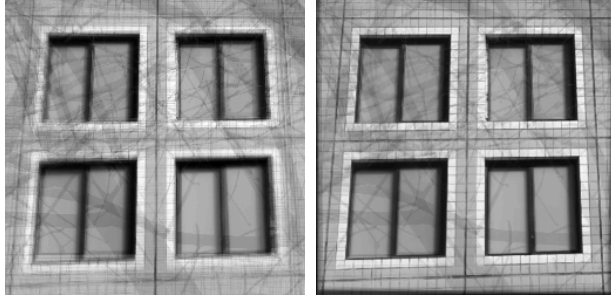


Figure 2. **Left panel:** superposition of the 10 observed images. **Right panel:** superposition of the 10 observed images after registration with the estimated parameters  $\theta_1^*, \dots, \theta_{10}^*$ .

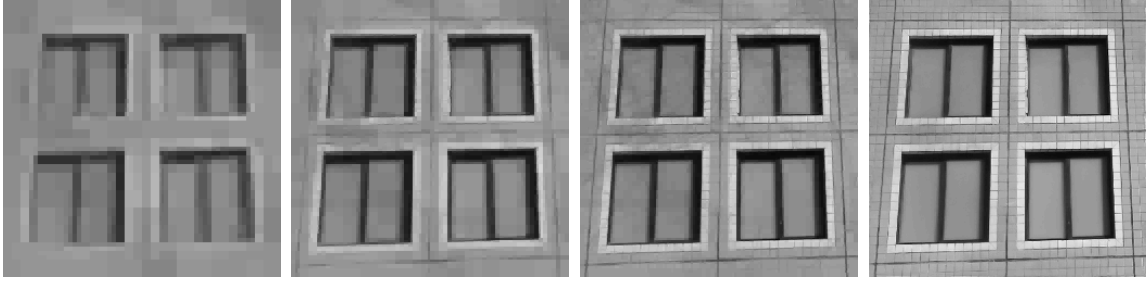


Figure 3. Sequence of estimated background images  $\mathbf{x}_0^{k+1}$  at  $k = 0$  (left),  $k = 10$  (middle left),  $k = 20$  (middle right), and at convergence (right).

In our experiment, we use  $l = 10$  occluded images of  $n = 256 \times 256$  pixels. The first 5 images are presented in the top panels of Fig. 1. We then run Algorithm 1 with  $\kappa = 100$ ,  $\lambda_\theta = 0.1$ ,  $\lambda_x^k = \max(0.9^k (20\kappa), 0.1)$ , and  $\mu = 10^{-10}$ . The matrix  $\Psi$  is build using the Haar wavelet basis, and  $f(\mathbf{x})$  is set to  $\|\Psi^\top \mathbf{x}\|_1$ . Finally, we assume that the transformations between images are well modeled by homographies and use model (3.16) for  $\tau_\theta$ . The parameters are constrained as follows:  $|1 - \theta_1|, |1 - \theta_5| \leq 0.2$ ,  $|\theta_2|, |\theta_4| \leq 0.2$ ,  $|\theta_3|, |\theta_6| \leq 20$ ,  $|\theta_7|, |\theta_8| \leq 0.01$ , with  $u_1$  and  $u_2$  belonging to  $\{-127, -126, \dots, 128\}$ .

The aligned background images  $S(\theta_j^*)\mathbf{x}_0^*$ , with  $j = 1, \dots, 5$ , and the estimated foreground images  $\mathbf{x}_1^*, \dots, \mathbf{x}_5^*$  are presented in Fig. 1. One can already notice that the separation background-foreground is very accurate. In particular, the background image does not contain any object initially occluding the scene. To have a better visualization of the quality of the estimated transformations parameters, we present in Fig. 2 the superposition of the 10 observed images before and after registration with the estimated parameters. One can easily remark that the registration is also very accurate.

In Fig. 3, we present the evolution of estimated background image  $\mathbf{x}_0^{k+1}$  at iterations  $k = 0$ ,  $k = 10$ ,  $k = 20$ ,  $k = 30$ , and at convergence. This sequence of images illustrate the fact that, as explained in Section 2.2, the images are reconstructed in a coarse-to-fine scales fashion.

For comparison, one can check in [24, 25] the performance reached by the low-rank and sparse approximation technique on the same dataset. Both methods provide very similar reconstructions. Note, however, that no guarantee about the convergence of the method described in [24, 25] is provided.

#### 4.2. Compressed sensing

For this second experiment, we study the performance of our algorithm in a compressed sensing setting, and compare it with other common algorithms.



Figure 4. **Top panels:** 5 initial images. **Middle-top panels:** reconstructed images with method (4.17). **Middle-bottom panels:** reconstructed images with method (4.18). **Bottom panels:** reconstructed images with the proposed method.

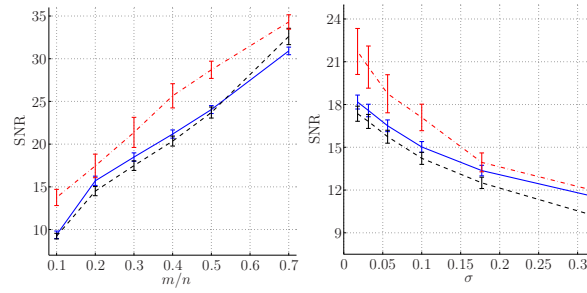


Figure 5. **Left panel:** Reconstruction SNR as a function of  $m/n$ . **Right panel:** Reconstruction SNR as a function of the noise level  $\sigma$ . All curves represents the mean SNR over the  $30 \times 5$  reconstructed images. The vertical lines shows the variation of the SNR at one standard deviation. The dashed black, continuous blue, and dot-dashed red curves represent the reconstruction SNR for method (4.17), (4.18), and our algorithm, respectively.



Figure 6. **Left panel:** superposition of the 5 initial images without registration. **Middle panel:** superposition of the 5 initial images after registration with the estimated parameters  $\theta_1^*, \dots, \theta_5^*$ . **Right panel:** estimated background image  $\mathbf{x}_0^*$ .

We use  $l = 5$  different images<sup>4</sup> of the same scene to generate 5 different measurement vectors  $\mathbf{y}_1, \dots, \mathbf{y}_l$ . These images, presented in the top panels of Fig. 4, contain  $n = 128 \times 128$  pixels. Note that parts of the scene are sometimes occluded. The measurement vectors are obtained using the spread spectrum technique of [28], which consists of pre-modulating each image with a random  $\pm 1$  sequence before randomly selecting  $m$  Fourier coefficients. Let  $\mathbf{c} = (c_1, \dots, c_n)^\top \in \mathbb{R}^n$  be a Rademacher sequence,  $\mathbf{F} \in \mathbb{C}^{n \times n}$  be the matrix that implements the 2D discrete Fourier transform, and  $\Omega \in \mathbb{N}^m$  be a set of  $m$  indices selected independently and uniformly at random in  $\{1, \dots, n\}$ . As we are dealing with images living in  $\mathbb{R}^n$ , we can restrict the selection of the Fourier coefficients to one half of the Fourier plane. We denote  $\tilde{\mathbf{F}} \in \mathbb{C}^{n/2 \times n}$  the restriction of  $\mathbf{F}$  to this half plane. The observation matrices satisfy

$$\mathbf{A}_1 = \dots = \mathbf{A}_l = \mathbf{A} = \begin{pmatrix} \tilde{\mathbf{F}}^r \\ \tilde{\mathbf{F}}^i \end{pmatrix}_\Omega \mathbf{C} \in \mathbb{R}^{m \times n},$$

where  $\tilde{\mathbf{F}}^r, \tilde{\mathbf{F}}^i \in \mathbb{R}^{n/2 \times n}$  are respectively the real and imaginary part of  $\tilde{\mathbf{F}}$ ,  $\mathbf{C} \in \mathbb{R}^{n \times n}$  is the diagonal matrix containing the sequence  $\mathbf{c}$ , and  $(\cdot)_\Omega$  restricts an  $n \times n$  matrix to its rows indexed by  $\Omega$ . Let us remark that, unlike here, one can choose a different subset  $\Omega$  for each matrix  $\mathbf{A}_i$  to introduce more innovations in the set of acquired measurements and to benefit even more of the correlation between the acquired signals when using a joint reconstruction technique like ours.

As in the last experiment, we assume that the transformations between images can be modeled by homographies and use model (3.16) for  $\tau_\theta$ . For the prior term  $f$ , we assume that the images are sparse in the Haar wavelet basis  $\mathbf{W} \in \mathbb{R}^{n \times n}$ . We build the block-diagonal matrix  $\Psi$  by repeating  $l + 1$  times the matrix  $\mathbf{W}$  on the diagonal, and set  $f(\mathbf{x}) = \|\Psi^\top \mathbf{x}\|_1$ . In all simulations, we use  $\lambda_\theta = 0.1$ ,  $\lambda_{\mathbf{x}}^k = \max(0.9^k(20\kappa), 0.1)$ ,  $\mu = 10^{-10}$ , and the following constraints apply on the transformations parameters:  $|1 - \theta_1|, |1 - \theta_5| \leq 0.2$ ,  $|\theta_2|, |\theta_4| \leq 0.2$ ,  $|\theta_3|, |\theta_6| \leq 20$ ,  $|\theta_7|, |\theta_8| \leq 0.01$ , with  $u_1$  and  $u_2$  belonging to  $\{-63, \dots, 64\}$ .

The first set of simulations study the performance of Algorithm 1 in the absence of noise for different number of measurements:  $m \in \{0.1n, 0.2n, 0.3n, 0.4n, 0.5n, 0.7n\}$ . The second set of simulations study the performance of Algorithm 1 at  $m = 0.3n$  for different noise levels. The noise vector  $\mathbf{n}$  follows an i.i.d zero-mean Gaussian distribution of standard deviation  $\sigma \in \{0.01, 0.0177, 0.0316, 0.0562, 0.1, 0.177, 0.316\}$ . The squared  $\ell_2$ -norm of the noise vector  $\mathbf{n}/\sigma$  follows a chi-square distribution with  $lm$  degrees of freedom and we compute the bound on the noise level  $\|\mathbf{n}\|_2^2 \leq \epsilon^2$  using the 99<sup>th</sup> percentile of this distribution. Then,  $\kappa$  is set to 100 in the noiseless case and is chosen so that  $\|\mathbf{A}(\theta^*)\mathbf{x}^* - \mathbf{y}\|_2 \approx \epsilon$  in the presence of noise<sup>5</sup>.

For comparison, we also reconstruct the images by solving the following convex optimization problems:

$$(4.17) \quad \min_{\mathbf{x}} \|\mathbf{W}^\top \mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{Y} - \mathbf{A}\mathbf{x}\|_F \leq \epsilon,$$

<sup>4</sup>castle-R20 dataset available at [cvlab.epfl.ch/~strecha/multiview/rawMVS.html](http://cvlab.epfl.ch/~strecha/multiview/rawMVS.html) [34].

<sup>5</sup>For each noise level, we use one simulation to estimate a value of  $\kappa$  such that, at converge of the algorithm,  $\|\mathbf{A}(\theta^*)\mathbf{x}^* - \mathbf{y}\|_2 \in [0.99\epsilon, 1.01\epsilon]$ . We then use the same value of  $\kappa$  for all subsequent simulations.

and

$$(4.18) \quad \min_{\mathbf{X}} \|\mathbf{W}^T \mathbf{X}\|_{2,1} \quad \text{subject to} \quad \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F \leq \epsilon,$$

where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_l) \in \mathbb{R}^{n \times l}$ ,  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_l) \in \mathbb{R}^{m \times l}$ ,  $\|\cdot\|_F$  is the Frobenius norm,  $\|\mathbf{X}\|_1 = \sum_{i=1}^n \sum_{j=1}^l |x_{ij}|$ , and  $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^n (\sum_{j=1}^l x_{ij}^2)^{1/2}$ , with  $x_{ij}$  the entry on the  $i^{\text{th}}$  line and  $j^{\text{th}}$  column of  $\mathbf{X}$ . In the first problem, the images  $\mathbf{x}_1, \dots, \mathbf{x}_l$  are reconstructed with a sparse prior in the Haar wavelet basis. This is an extension of the Basis Pursuit problem, advocated by the compressed sensing theory [12], to multiple signals. In the second problem, the prior term favors reconstructions which have a similar sparsity pattern in the Haar wavelet basis [5], therefore imposing some correlation between images.

Fig. 5 shows the quality of the reconstructed images in term of SNR. Let  $(\mathbf{x}^*, \boldsymbol{\theta}^*) \in \mathbb{R}^{(l+1)n} \times \mathbb{R}^{ln}$  be the images and transformation parameters recovered with our algorithm and  $\mathbf{x} \in \mathbb{R}^{ln}$  be the initial  $l$  images. The reconstruction SNR satisfies  $-20 \log_{10}(\|\mathbf{S}(\boldsymbol{\theta}_j^*)\mathbf{x}_0^* + \mathbf{x}_j^* - \mathbf{x}_j\|_2 / \|\mathbf{x}_j\|_2)$ . For the two other methods (4.17) and (4.18), the reconstruction SNR satisfies  $-20 \log_{10}(\|\mathbf{x}_j^* - \mathbf{x}_j\|_2 / \|\mathbf{x}_j\|_2)$  where  $\mathbf{X}^* = (\mathbf{x}_1^*, \dots, \mathbf{x}_l^*) \in \mathbb{R}^{n \times l}$  are the reconstructed images with these methods. The first graph of Fig. 5 shows the reconstructions SNR as a function of  $m/n$  in the absence of noise, and the second graph shows the reconstructions SNR as a function of the noise level  $\sigma$  for  $m/n = 0.3$ . In each case, we perform 30 simulations with independent noise realization and choice of  $\Omega$ . The curves show the mean SNR over the  $30 \times 5$  reconstructed images. The vertical lines indicate the variation of the SNR at one standard deviation. One can deduce from these curves that the best reconstructions are obtained with our method.

In Fig. 4, we present the ground truth images as well as reconstructions obtained with the three different methods for  $m/n = 0.3$  in the absence of noise. One can remark that the reconstructions obtained with our algorithm exhibits much finer details. Fig. 6 shows the estimated background image, next to the superposed initial images before and after registration with the estimated parameters. One can also note that the estimated background image is free of occlusions, and that the initial images are better aligned after registration with the estimated parameters.

### 4.3. Super-resolution

For this last experiment, we test our algorithm for super-resolution from multiple frames, as considered in, e.g., [16, 41].

We use  $l = 16$  low-resolution images<sup>6</sup> with  $m = 64 \times 64$  pixels extracted from a video sequence and try to increase the resolution by 2 in both directions, i.e.,  $n = 128 \times 128$ .

To increase the resolution of these images, we use the measurement model (1.2) where the sensing matrices  $\mathbf{A}_j$ , with  $j = 1, \dots, l$ , model a blurring and downsampling operator that averages all the pixels in a rectangular window of  $2 \times 2$  pixels and uniformly downsamples the resulting blurred image by a factor of 2 in both dimensions. The sensing matrices are identical for all  $j$ . The vectors  $\mathbf{y}_1, \dots, \mathbf{y}_l$  are initialized with the initial low-resolution images. The transformations between images are assumed to be affine, and we use model (3.15) for  $\tau_{\boldsymbol{\theta}}$ . Then, we assume that the high-resolution images  $\mathbf{x}_0, \dots, \mathbf{x}_l$  are sparse in the gradient domain and set  $f(\mathbf{x}) = \sum_{j=0}^l g(\mathbf{x}_j)$  in (2.4), where  $g: \mathbb{R}^n \rightarrow \mathbb{R}$  is the anisotropic Total Variation (TV) norm. Let us recall that the vector  $\mathbf{x}_j$  contains the samples  $x_j(\mathbf{u}_k)$ , with  $k = 1, \dots, n$  and  $\mathbf{u}_k = (u_1^k, u_2^k)$  (see Section 1.1). The anisotropic TV norm  $g$  satisfies

$$g(\mathbf{x}_j) = \sum_{k=1}^n |\nabla_1 x_j(\mathbf{u}_k)| + |\nabla_2 x_j(\mathbf{u}_k)|,$$

where

$$\begin{cases} \nabla_1 x_j(\mathbf{u}_k) = x_j(u_1^{k+1}, u_2^k) - x_j(u_1^k, u_2^k), \\ \nabla_2 x_j(\mathbf{u}_k) = x_j(u_1^k, u_2^{k+1}) - x_j(u_1^k, u_2^k), \end{cases}$$

<sup>6</sup>Dataset available at <http://users.soe.ucsc.edu/~milanfar/software/sr-datasets.html> (Credit: Peyman Milanfar).

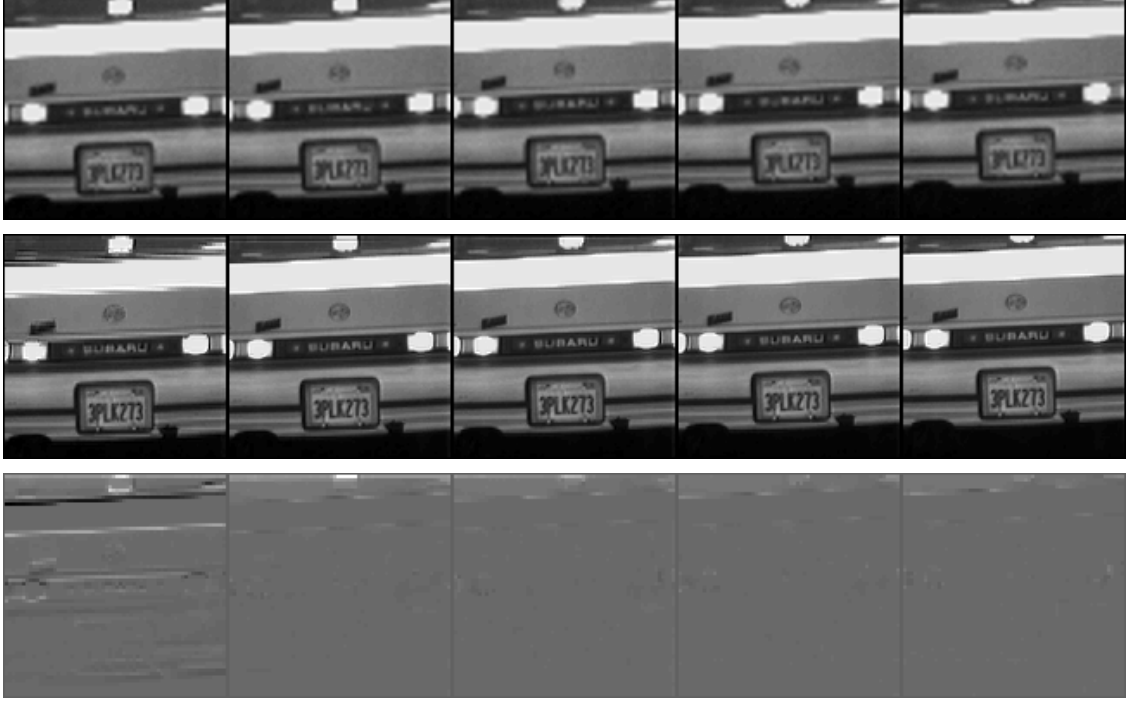


Figure 7. **Top panels:** first 5 input images upsampled independently by a factor 2 in both dimension using a bicubic interpolation. **Middle panels:** first 5 reconstructed high resolution images  $S(\theta_j^*)\mathbf{x}_0^* + \mathbf{x}_j^*$ , with  $j = 1, \dots, 5$ , with our algorithm. **Bottom panels:** first 5 reconstructed foreground images  $\mathbf{x}_1^*, \dots, \mathbf{x}_5^*$ .



Figure 8. **Left panel:** superposition of the 16 input images upsampled independently by a factor 2 in both dimension using a bicubic interpolation. **Middle panel:** superposition of the same 16 images after registration with the estimated parameters  $\theta_1^*, \dots, \theta_{16}^*$ . **Right panel:** estimated background image  $\mathbf{x}_0^*$  at high-resolution.

assuming that  $x(u_1^{n+1}, \cdot) = x(\cdot, u_2^{n+1}) = 0$ . Note that  $g$  and, consequently,  $f$  are semi-algebraic. Finally, we run Algorithm 1 to minimize (2.4) with  $\kappa = 100$ ,  $\lambda_\theta = 0.1$ ,  $\lambda_x^k = \max(0.9^k(20\kappa), 0.1)$ , and  $\mu = 10^{-10}$ . We also constrain the transformation parameters as follows:  $|1 - \theta_1|, |1 - \theta_5| \leq 0.2$ ,  $|\theta_2|, |\theta_4| \leq 0.2$ ,  $|\theta_3|, |\theta_6| \leq 64$ , with  $u_1$  and  $u_2$  belonging to  $\{-63, \dots, 64\}$ .

In Fig. 7, we present the first 5 super-resolved images with 2 different techniques. The top panels show the image upsampled independently using a bicubic spline interpolation, and the middle panels show the super-resolved images using our method, which exploits the inter-correlation between images. Note that the brand of the car and license plate number become readable with our reconstructions. We also present, in the bottom panels, the first 5 reconstructed foreground images which, here, essentially correct slight registration errors for the first reconstructed image.

In Fig. 8, we present the estimated background image  $\mathbf{x}_0^*$  alone, and the input images superposed before and after registration. One can notice that the images are correctly aligned after registration.



## 5. Conclusion

We presented a method to solve multi-view imaging problems where one has to reconstruct several images of a scene from only a few linear observations made at different viewpoints. Each observed image is modeled as the sum of a geometrically transformed background image and of a foreground image, modeling possible occlusions. We considered here global geometric transformations represented by few parameters, such as translations, affine transformations, or homographies. The proposed reconstruction method jointly estimates the images (background and foreground) and the transformations parameters by minimizing a non-convex functional. We studied the convergence of the proposed algorithm and showed that the generated sequence of estimates converges to a critical point of the non-convex functional for commonly used priors and transformation models.

An obvious limitation of our technique arises when the geometric transformations of the background image between two different observations are not global but local. This situation for example occurs when, unlike the images used in this paper, the depth of the observed scene is not constant. In this case, the geometric transformations differ from one pixel location to another. However, the following two properties of the proposed method may help to compensate for such an issue. First, one should note that the presence of foreground images in the measurement model can render our method robust to a small transformation model mismatch by considering as “occlusions” the parts of the scene that cannot be aligned with the assumed model. Second, let us also remark that the requirements of Theorem 1 hold for a large class of transformation models  $\tau_\theta$  and leave us the possibility to choose more general type of transformations. For example, we could approximate elastic transformations using a polynomial of the spatial coordinates (see, e.g., [20]) and estimate the coefficients of this polynomial using the proposed algorithm.

Finally, we would like to conclude by highlighting the potential interest of the proposed technique in free breathing coronary magnetic resonance imaging (MRI) [33]. One of the major challenges in this application is to handle respiratory motion properly. The low acquisition speed in MRI forces researchers to invent new techniques to compensate for the inevitable motions occurring during the acquisition. To avoid motions due to heart contractions, an ECG signal is acquired to ensure that the Fourier measurements are always taken at the same instant in the heart cycle. A few measurements are then taken at each cycle. As the number of measurements acquired during one cycle is too small to accurately reconstruct an image of the heart, one should combine the measurements of several cycles to gather enough information. It is thus mandatory to properly compensate for the respiratory motion occurring between heart beats in the reconstruction method. Let us remark that simple geometric transformations, such as translations, are already sufficient to reach good image quality [11]. Consequently, as the technique developed in this paper automatically compensates for such geometric transformations, it is a promising method for such an application.

## Appendix A

To establish the third point of Theorem 1, we need the following theorem established by Attouch *et al.* in [4].

**Theorem 2** (Theorem 6.2, [4]). *Let  $p$  be a positive integer larger than 2,  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$  be a vector in  $\mathcal{H} = \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$ , and  $g: \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  a function of the form  $g(\mathbf{x}) = Q(\mathbf{x}) + \sum_{i=1}^p g_i(\mathbf{x}_i)$ , where  $Q: \mathcal{H} \rightarrow \mathbb{R}$  is a  $C^1$  function with locally Lipschitz continuous gradient and  $g_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous function,  $i = 1, \dots, p$ . Assume that  $g$  is bounded from below and satisfies the Kurdyka-Łojasiewicz property. Let  $(\mathbf{x}^k)_{k \in \mathbb{N}}$  and  $(\boldsymbol{\nu}^k)_{k \in \mathbb{N}}$  be sequences such that*

$$\begin{aligned}
 & g_i(\mathbf{x}_i^{k+1}) + Q(\mathbf{x}_1^{k+1}, \dots, \mathbf{x}_{i-1}^{k+1}, \mathbf{x}_i^{k+1}, \dots, \mathbf{x}_p^k) + a \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|_2^2 \\
 & \leq g_i(\mathbf{x}_i^k) + Q(\mathbf{x}_1^{k+1}, \dots, \mathbf{x}_{i-1}^{k+1}, \mathbf{x}_i^k, \dots, \mathbf{x}_p^k), \\
 & \boldsymbol{\nu}_i^{k+1} \in \partial g_i(\mathbf{x}_i^{k+1}), \\
 & \|\boldsymbol{\nu}_i^{k+1} + \partial_{\mathbf{x}_i} Q(\mathbf{x}_1^{k+1}, \dots, \mathbf{x}_{i-1}^{k+1}, \mathbf{x}_i^{k+1}, \mathbf{x}_{i+1}^k, \dots, \mathbf{x}_p^k)\|_2 \\
 & \leq b \|\mathbf{x}_i^{k+1} - \mathbf{x}_i^k\|_2,
 \end{aligned}
 \tag{A.19}$$

where  $i$  ranges over  $\{1, \dots, p\}$  and  $a, b > 0$ . Then, if the sequence  $(\mathbf{x}^k)_{k \in \mathbb{N}}$  is bounded, it converges to some critical point  $\mathbf{x}^*$  of  $g$ .

The objective function  $L$  defined in (2.4) is equal to

$$L(\mathbf{x}, \boldsymbol{\theta}) = \kappa Q(\mathbf{x}, \boldsymbol{\theta}) + f(\mathbf{x}) + i_{\Theta}(\boldsymbol{\theta}),$$

where  $Q: \mathbb{R}^{(l+1)n} \times \mathbb{R}^{lq} \rightarrow \mathbb{R}$  satisfies  $Q(\mathbf{x}, \boldsymbol{\theta}) = \|\mathbf{A}(\boldsymbol{\theta})\mathbf{x} - \mathbf{y}\|_2^2$ . The objective function has the form required by Theorem 2. Furthermore,  $f$  and  $i_{\Theta}$  are proper lower semicontinuous functions, and, using the assumptions of Theorem 1,  $Q$  is twice continuously differentiable and thus have locally Lipschitz continuous gradient. We also note that  $L$  is bounded below and satisfies the Kurdyca-Łojasiewicz property. To prove the third point of Theorem 1, it remains to show that the sequence generated by Algorithm 1 is bounded and satisfies conditions (A.19).

First, due to the constraints that applies on  $\boldsymbol{\theta}$ , it is obvious that the sequence  $(\boldsymbol{\theta}^k)_{k \in \mathbb{N}}$  is bounded. Therefore, assuming that  $(\mathbf{x}^k)_{k \in \mathbb{N}}$  is bounded, as in the third point of Theorem 1, is enough to ensure the boundedness condition of the entire sequence of estimates.

Second,  $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2$  tends to 0 as  $k \rightarrow +\infty$ , as established in the second point of Theorem 1. Therefore there exists  $k_0 \in \mathbb{N}$  such that for all  $k \geq k_0$ ,  $h_{\mu}(\Psi^{\top}(\mathbf{x}^{k+1} - \mathbf{x}^k)) = \|\Psi^{\top}(\mathbf{x}^{k+1} - \mathbf{x}^k)\|_2^2 / (2\mu) = \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 / (2\mu)$ , as  $\Psi\Psi^{\top} = \mathbf{I}_{(l+1)n}$ . Let  $k$  be larger than  $k_0$  in the following. Inequality (2.6) shows that

$$f(\mathbf{x}^{k+1}) + \kappa Q(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k) + \frac{\lambda}{4\mu} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2^2 \leq f(\mathbf{x}^k) + \kappa Q(\mathbf{x}^k, \boldsymbol{\theta}^k).$$

Furthermore, the first order optimality condition of Problem (2.5) shows that there exists  $\boldsymbol{\nu}_x^{k+1} \in \partial f(\mathbf{x}^{k+1})$  such that

$$\boldsymbol{\nu}_x^{k+1} + \kappa \partial_{\mathbf{x}} Q(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k) + \frac{\lambda \mathbf{x}}{2\mu} (\mathbf{x}^{k+1} - \mathbf{x}^k) = 0,$$

which implies,

$$\|\boldsymbol{\nu}_x^{k+1} + \kappa \partial_{\mathbf{x}} Q(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k)\|_2 \leq \frac{\bar{\lambda}}{2\mu} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2.$$

Finally, we deduce from equation (2.11) that

$$i_{\Theta}(\boldsymbol{\theta}^{k+1}) + \kappa Q(\mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1}) + \frac{\kappa \lambda_{\boldsymbol{\theta}}}{2} \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2^2 \leq i_{\Theta}(\boldsymbol{\theta}^k) + \kappa Q(\mathbf{x}^{k+1}, \boldsymbol{\theta}^k).$$

And, using the first order optimality condition of the problems (2.9), we conclude that there exists  $\boldsymbol{\nu}_{\boldsymbol{\theta}_j}^{k+1} \in \partial i_{\Theta_j}(\boldsymbol{\theta}_j^{k+1})$  such that

$$\boldsymbol{\nu}_{\boldsymbol{\theta}_j}^{k+1} + \kappa \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^k) + \kappa [\mathbf{H}(\boldsymbol{\theta}_j^k) + 2^i \lambda_{\boldsymbol{\theta}} \mathbf{I}_q](\boldsymbol{\theta}_j^{k+1} - \boldsymbol{\theta}_j^k) = 0,$$

for all  $j = 1, \dots, l$ . Let  $\Gamma$  denote the Lipschitz constant of  $\nabla Q$  on a product of balls  $\mathcal{B}_{\mathbf{x}} \times \mathcal{B}_{\boldsymbol{\theta}}$  containing the sequence  $(\mathbf{x}_k, \boldsymbol{\theta}_k)_{k \in \mathbb{N}}$ . By construction of  $i$  in Algorithm 1, we necessarily have  $2^i \lambda_{\boldsymbol{\theta}} \leq \eta = \max(2\Gamma, \lambda_{\boldsymbol{\theta}})$ . Let us also remark that, on the same products of balls, there exists  $\Lambda$  such that the singular values of the matrices  $\mathbf{H}(\boldsymbol{\theta}_j^k)$ ,  $j = 1, \dots, l$ , are bounded above by  $\Lambda$ . Therefore,

$$\begin{aligned} & \|\boldsymbol{\nu}_{\boldsymbol{\theta}_j}^{k+1} + \kappa \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^{k+1})\|_2 \\ &= \|\boldsymbol{\nu}_{\boldsymbol{\theta}_j}^{k+1} + \kappa \nabla Q_j^k(\boldsymbol{\theta}_j^k) + \kappa \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^{k+1}) - \kappa \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^k)\|_2 \\ &\leq \|\boldsymbol{\nu}_{\boldsymbol{\theta}_j}^{k+1} + \kappa \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^k)\|_2 + \kappa \|\nabla Q_j^{k+1}(\boldsymbol{\theta}_j^{k+1}) - \nabla Q_j^{k+1}(\boldsymbol{\theta}_j^k)\|_2 \\ &\leq \kappa(\Lambda + \eta) \|\boldsymbol{\theta}_j^{k+1} - \boldsymbol{\theta}_j^k\|_2 + \kappa \Gamma \|\boldsymbol{\theta}_j^{k+1} - \boldsymbol{\theta}_j^k\|_2 \\ &= \kappa(\Lambda + \eta + \Gamma) \|\boldsymbol{\theta}_j^{k+1} - \boldsymbol{\theta}_j^k\|_2, \end{aligned}$$

which yields,

$$\|\boldsymbol{\nu}_{\boldsymbol{\theta}}^{k+1} + \kappa \partial_{\boldsymbol{\theta}} Q(\mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1})\|_2 \leq \kappa (\Lambda + \eta + \Gamma) \|\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\|_2,$$

with  $\nu_{\theta}^{k+1} = ((\nu_{\theta_1}^{k+1})^\top, \dots, (\nu_{\theta_l}^{k+1})^\top)^\top \in \partial i_{\Theta}(\theta^{k+1})$ . The sequence  $(x^k, \theta^k)_{k \geq k_0}$  thus satisfies conditions (A.19) with  $a = \min(\underline{\lambda}/(4\mu), \kappa\lambda_{\theta}/2)$  and  $b = \max(\bar{\lambda}/(2\mu), \kappa(\Lambda + \eta + \Gamma))$ . This terminates the proof of the third point of Theorem 1.

## References

- [1] AFSHARI, H., JACQUES, L., BAGNATO, L., SCHMID, A., VANDERGHEYNST, P., AND LEBLEBICI, Y. The panoptic camera: A plenoptic sensor with real-time omnidirectional capability. *J. Signal Process. Syst. Signal Image Video Technol. accepted* (2012).
- [2] ASIF, M. S., HAMILTON, L., BRUMMER, M., AND ROMBERG, J. Motion-adaptive spatio-temporal regularization (master) for accelerated dynamic mri. *Magn. Reson. Med., submitted* (2012).
- [3] ATTOUCH, H., BOLTE, J., REDONT, P., AND SOUBEYRAN, A. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-lojasiewicz inequality. *Math. Oper. Res.* 35, 2 (2010), 438–457.
- [4] ATTOUCH, H., BOLTE, J., AND SVAITER, B. F. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. *Math. Program.* (2011).
- [5] BACH, F., JENATTON, R., MAIRAL, J., AND OBOZINSKI, G. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning* 4, 1 (2012), 1–106.
- [6] BARON, D., WAKIN, M. B., DUARTE, M. F., SARVOTHAM, S., AND BARANIUK, R. G. Distributed compressed sensing. Tech. Rep. ECE-0612, Electrical and Computer Engineering Department, Rice University, 2006.
- [7] BAUSCHKE, H. H., AND COMBETTES, P. L. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics. Springer New York, 2011.
- [8] BECK, A., AND TEOULLE, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2, 1 (2009), 183–202.
- [9] BERG, E. V., FRIEDLANDER, M. P., HENNENFENT, G., HERRMANN, F., SAAB, R., AND YILMAZ, Ö. Sparco: A testing framework for sparse reconstruction. Tech. Rep. TR-2007-20, Dept. Computer Science, University of British Columbia, Vancouver, October 2007.
- [10] BERTSEKAS, D. P. Projected newton methods for optimization problems with simple constraints. *SIAM J. Control and Optimization* 20, 2 (1982), 221–246.
- [11] BONANNO, G., PUY, G., WIAUX, Y., VAN HEESWIJK, R., AND STUBER, M. About the performance of multi-dimensional radial self-navigation incorporating compressed sensing for free-breathing coronary mri. In *International Society for Magnetic Resonance in Medicine (ISMRM) conference* (Melbourne, 2012).
- [12] CANDÈS, E. J. Compressive sampling. In *Proc. International Congr. of Mathematicians* (Madrid, Spain, 2006), vol. 3, pp. 1433–1452.
- [13] CHEN, X., AND FROSSARD, P. Joint reconstruction of compressed multi-view images. In *Proc. IEEE Conf. Acoustics, Speech and Signal Processing* (2009), pp. 1005 – 1008.
- [14] DAI, Q., FU, C., JI, Y., AND ZHANG, Y. Robust joint reconstruction in compressed multi-view imaging. In *Proc. Picture Coding Symp.* (Poland, May 2012), pp. 13–16.
- [15] DUARTE, M. F., SARVOTHAM, S., WAKIN, M. B., BARON, D., AND BARANIUK, R. G. Joint sparsity models for distributed compressed sensing. In *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations* (2005).
- [16] FARSIU, S., ROBINSON, D., ELAD, M., AND MILANFAR, P. Fast and robust multi-frame super-resolution. *IEEE Trans. Image Process.* 12, 10 (2004), 1327–1344.
- [17] FU, C., JI, Y., AND DAI, Q. Compressed multi-view imaging with joint reconstruction. In *Proc. Data Compression Conf.* (2011), p. 448.
- [18] KEYS, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoustics, Speech, Signal Process.* 29, 6 (1981).
- [19] KUBOTA, A., SMOLIC, A., MAGNOR, M., TANIMOTO, M., CHEN, T., AND ZHANG, C. Multiview imaging and 3d tv. *IEEE Signal Process. Mag.* 24, 6 (2007), 10–21.
- [20] KYBIC, J., AND UNSER, M. Fast parametric elastic image registration. *IEEE Trans. Image Process.* 12, 11 (2003), 1427–1442.
- [21] LI, X., WEI, Z., AND XIAO, L. Compressed sensing joint reconstruction for multi-view images. *Electronics Lett.* 46, 23 (2010), 1548 – 1550.
- [22] MAMAGHANIAN, H., KHALED, N., ALONSO, D. A., AND VANDERGHEYNST, P. Design and exploration of low-power analog to information conversion based on compressed sensing. *IEEE J. Emerg. Sel. Topics in Circuits Syst.* (2012).
- [23] PARK, J. Y., AND WAKIN, M. B. A geometric approach to multi-view compressive imaging. *EURASIP J. Adv. Signal Process.* 2012, 37 (2012).
- [24] PENG, Y., GANESH, A., WRIGHT, J., XU, W., AND MA, Y. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *IEEE Conf. Computer Vision and Pattern Recognition* (2010), pp. 763 – 770.

- [25] PENG, Y., GANESH, A., WRIGHT, J., XU, W., AND MA, Y. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 11 (2012), 2233–2246.
- [26] PROTTER, M., ELAD, M., TAKEDA, H., AND MILANFAR, P. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Trans. Image Process.* 18, 1 (2009), 36–51.
- [27] PUY, G., AND VANDERGHEYNST, P. Robust joint reconstruction of misaligned images using semi-parametric dictionaries. In *ICML Workshop on Sparsity, Dictionaries and Projections in Machine Learning and Signal Processing* (Edinburgh, Scotland, 2012).
- [28] PUY, G., VANDERGHEYNST, P., GRIBONVAL, R., AND WIAUX, Y. Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques. *EURASIP J. Adv. Signal Process.* 2012, 6 (2012).
- [29] RUDIN, L. I., OSHER, S., AND FATEMI, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60, 1-4 (1992), 259–268.
- [30] SANKARANARAYANAN, A. C., TURAGA, P. K., BARANIUK, R. G., AND CHELLAPPA, R. Compressive acquisition of dynamic scenes. In *Computer Vision - ECCV 2010* (2010), vol. 6311, pp. 129–142.
- [31] SCHMIDT, M., KIM, D., AND SRA, S. *Optimization for Machine Learning*. MIT press, 2011, ch. Projected Newton-type methods in machine learning.
- [32] SHEN, H., ZHANG, L., HUANG, B., AND LI, P. A map approach for joint motion estimation, segmentation, and super resolution. *IEEE Trans. Image Process.* 16, 2 (2007), 479–490.
- [33] STEHNING, C., BÖRNERT, P., NEHRKE, K., EGGERS, H., AND STUBER, M. Free-breathing whole-heart coronary mra with 3d radial ssfp and self-navigated image reconstruction. *Magn. Reson. Med.* 54, 2 (2005), 476–480.
- [34] STRECHA, C., VON HANSEN, W., GOOL, L. V., FUA, P., AND THOENNESSEN, U. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *IEEE Conf. Computer Vision and Pattern Recognition* (2008), pp. 1 – 8.
- [35] THIRUMALAI, V., AND FROSSARD, P. Joint reconstruction of multi-view compressed images. *IEEE Trans. Image Process.*, submitted (2012).
- [36] TOLA, E., STRECHA, C., AND FUA, P. Efficient large-scale multi-view stereo for ultra high-resolution image sets. *Mach. Vis. Appl.* 23, 5 (2012), 903–920.
- [37] TROCAN, M., MAUGEY, T., FOWLER, J. E., AND PESQUET-POPESCU, B. Disparity-compensated compressed-sensing reconstruction for multiview images. In *Proc. IEEE Conf. Multimedia and Expo* (2010), pp. 1225 – 1229.
- [38] TROCAN, M., MAUGEY, T., TRAMEL, E. W., FOWLER, J. E., AND PESQUET-POPESCU, B. Compressed sensing of multiview images using disparity compensation. In *Proc. IEEE Conf. Image Processing* (2010), pp. 3345 – 3348.
- [39] TSENG, P. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* 109, 3 (2001), 475–494.
- [40] UNSER, M. Sampling—50 years after shannon. *Proc. IEEE* 88, 4 (2000), 569– 587.
- [41] VANDEWALLE, P., SÜSSTRUNK, S., AND VERTELLI, M. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP J. Applied Signal Process.* 2006 (2006), 71459.
- [42] WAKIN, M. B. A manifold lifting algorithm for multi-view compressive imaging. In *Proc. Picture Coding Symp.* (2009), pp. 1 – 4.
- [43] WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. High performance imaging using large camera arrays. *ACM Trans. Graphics* 24, 3 (2005), 765–776.